基于深度强化学习的移动边缘计算资源分配策略

冯北鹏　黄昱泽*　曹宇慧　郭珍珍

1)(重庆交通大学信息科学与工程学院 重庆市　400074)

摘要：云计算可解决移动设备计算资源不足的问题，但无法满足低时延的服务需求，边缘计算作为云计算技术的延伸，可通过增强边缘网络计算能力从而为用户提供低时延高质量服务。边缘计算中，需要将服务部署于资源受限的边缘服务器，并根据需求合理分配计算资源，以提高边缘服务器资源利用率，为此，本文提出了一种基于深度强化学习的服务资源分配方法，利用反正切函数两次映射建立计算资源分配函数，并实现分配比例的动态调整，最后基于真实数据集进行仿真实验，实验结果表明，本文提出的方法能够在保证低时延的情况下，合理分配计算资源。

关键词：移动边缘计算；深度强化学习；资源分配

分类号：TP311.5

Strategy of Resource Allocation Based on Deep Reinforcement Learning in Mobile Edge Computing

FENG Bei-peng　HUANG Yu-ze　CAO Yu-hui　GUO Zhen-Zhen

1)(School of Information Science and Engineering,Chongqing Jiaotong University,Chongqing 400074,China)

Abstract：Cloud computing can solve the problem of insufficient computing resources of mobile devices, but cannot meet the service demand of low latency. As an extension of cloud computing technology, edge computing can provide users with low latency and high-quality services by enhancing the computing capability of edge networks. In edge computing, services need to be deployed to resource constrained edge servers, and computing resources need to be allocated reasonably according to the needs to improve the utilization of edge server resources. For this reason, this paper proposes a service resource allocation method based on deep reinforcement learning, uses arctangent function twice to map to establish computing resource allocation function, and realizes dynamic adjustment of allocation ratio. Finally, simulation experiments are conducted based on real data sets, The experimental results show that the proposed method can reasonably allocate computing resources with low latency.

Key words：Mobile edge computing; Deep reinforcement learning; resource allocation

1 引言

今天是万物互联的时代，各种移动设备纷纷踏上 5G 网络的高速列车加入到互联网，导致网络上的数据呈指数级增长[1]，给网络负载带来巨大的挑战，且随着时延敏感型应用的开发和投入，对硬件的资源不断提出新的要求[2]。在此背景下，移动边缘计算作为一种新的计算模式被提出[3]，它可以在靠近用户附近的位置给无线接入网提供运算支持，从而使得更多应用的普及成为可能。但是也应该注意到，边缘服务器作为边缘计算的主要处理单元它的资源也是十分稀缺的，尤其是分布式的特征使得单个边缘服务器不可能承载多个服务，所以当用户请求传至边缘网络时通常是由部署有该服务的服务器完成[4]，每个服务器对其所承载的

本文系重庆市自然科学基金项目(CSTB2022NSCQ-MSX0368); 重庆市教育委员会科学技术研究计划青年项目(KJQN202200702,KJQN201900708)研究成果之一。

服务分配资源的大小就会直接影响到服务响应时延。

服务资源分配的优化目标通常为为减小能耗和平均响应时延[5-8]:在 5G 异构网络中提出一种以节能为目的的无线电资源分配的卸载方案，该方法能够适应复杂多变的网络环境，基于数据传输单项能耗模型进行优化，在能耗方面有较好的表现，但对于服务时延的优化略显不足，用户体验不好；有注重用户体验，在资源有限的条件下优化服务响应时延的边缘服务器资源的分配方法，但算法本身复杂度较高，对应用场景的要求较为苛刻，并不适用多变的移动边缘计算环境；对云端-边缘端-终端三层架构进行全局优化的资源分配方法通常收到多方面条件的限制，服务响应时延满意度难免差强人意。这其中，深度强化学习算法以其更接近人类的思维方式，智能体通过与环境交互不断试错和训练的方式得出最优行为策略，在解决资源分配问题上，不仅能够充分发挥算法解决全局问题的优势，还能适应不同的移动边缘计算环境，是解决此类问题的有效方式。

因此，针对移动边缘计算网络资源有限又要满足用户体验的背景下，提出了基于深度强化学习的计算资源分配方法，该方法先获得分配资源后的反馈，经过反正切函数两次映射动态调整下一次的资源分配比例，优化服务平均响应时延。

2 系统建模

2.1 边缘服务器与服务模型

如图 1 所示，为移动边缘计算场景，在该场景中由边缘服务器组成的边缘网络为终端用户提供稳定的运算服务，云端服务器充当资源调度中心，将 DQN（Deep Q-Learning）训练以后的资源分配决策下发到各边缘服务器。

边缘服务器集合 X_x 可以用一组四元参数来表示:

$$X_x = \left\{ Z_x, q_x^y, h_x^y\left(t\right), n_x^y\left(t\right) \right\} \tag{1}$$

其中四个变量分别表示边缘服务器 x 的计算资源、是否部署有服务 y 、在 t 时隙分配给服务 y 的计算资源所占边缘服务器 x 总资源的比例、在 t 时隙接收到的服务请求 y 的数量，q_x^y 为 0 或 1，为 1 时表示边缘服务器 x 部署有服务 y ，否则没有部署服务 y 。

服务 Y_y 用一组二元参数来表示：

$$Y_y = \left\{ g_y, u_y \right\} \tag{2}$$

g_y 、u_y 表示服务 y 请求数据转发量和处理所需的计算资源。

边缘服务器的计算资源有限，任一边缘服务器 x 所分配的计算资源比例不会超过自身：

$$\sum\nolimits_{x=1}^{X} h_x^y\left(t\right) \le 1, \forall t \tag{3}$$

2.2 服务处理时延模型

由于通常服务处理结果的传回数据量较小，因此服务处理过程中涉及到的时延主要包括请求转发时延和边缘服务器计算时延。

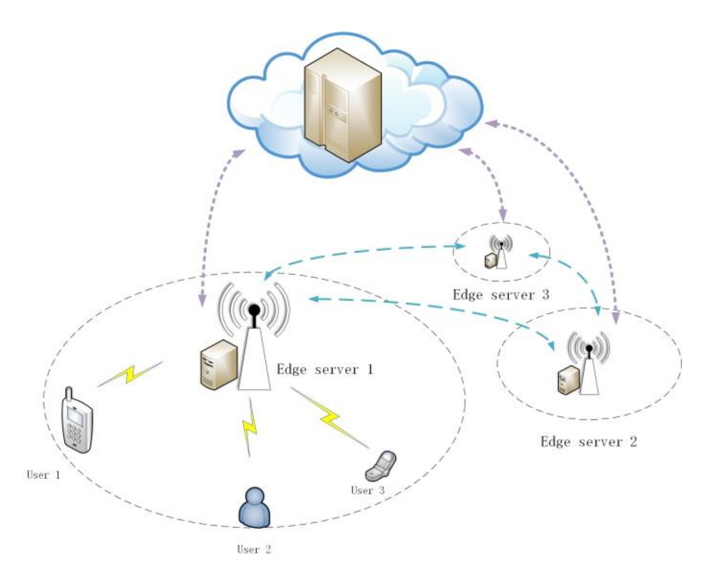


图 1 移动边缘计算架构图

当终端用户向边缘服务器 x 发出服务请求 y 以后，若边缘服务器部署有服务 y，则由边缘服务器计算处理后将结果传回，该过程包括终端与边缘服务器之间的数据传输时延和边缘服务器的计算时延，可以表示为：

$$f_y^1(t)=n_x^y(t)\left(\frac{g_y}{b_{edge}}+\frac{u_y}{h_x^y(t)Z_x}\right) \tag{4}$$

其中 $n_x^y(t)$ 为边缘服务器 x 在 t 时隙内接收到的服务请求 y 的数量，b_{edge} 为数据传输速率。当边缘服务器 x 接收到终端用户发出的服务请求 y 以后，若接收服务器没有部署该服务，则需要将该服务请求转发至其他部署有该服务的边缘服务器，因此，在时隙 t 需要转发的服务请求 y 的数量可以表示为：

$$N_y(t)=\sum_{x=1}^{X}\left(1-q_x^y\right)n_x^y(t) \tag{5}$$

则由服务请求产生的数据传输时延可以表示为：

$$f_y^2(t)=\frac{N_y(t)g_y}{b_{edge}} \tag{6}$$

因此可以得到处理服务请求 y 的总时延为：

$$F_y^{sum}(t)=\sum_{x=1}^{X}f_y^1(t)+f_y^2(t) \tag{7}$$

我们的目标是通过计算资源的分配实现缩短服务平均响应时延，问题可以表示为：

$$C=\min\frac{1}{T}\sum_{t=1}^{T}\sum_{y=1}^{Y}F_y^{sum}(t) \tag{8}$$

$$s.t.\quad \sum_{y=1}^{Y}h_x^y(t)\le 1,\forall t$$

$$h_x^y(t)\in[0,1],x\in X,y\in Y. \tag{9}$$

其中，公式（9）限制了可分配计算资源的量。为了求出目标函数的解，我们用深度强化学习算法训练得出。

2.3 基于深度强化学习的服务计算资源分配算法

深度强化学习建模通常是基于马尔可夫决策过程进行的，在本场景中主要包括以下几个要素：环境、状态、动作和奖励，因此，首先需要构建状态空间 S 、动作空间 A 和奖励函数 R 。

$$S_x(t)=\{n_x^y(t),Z_x,H(t)\} \tag{10}$$

动作空间主要为每个边缘服务器分配给相应服务 y 的计算资源比例，在我们的算法中，边缘服务器 x 分配给服务 y 计算资源量不是一次性决定的，而是通过一次次小比例的分配得到最终分配比例。值得注意的是，算法需要通过每次分配以后的反馈来动态调整下一次的分配大小，因此，在时隙 t 内，对于边缘服务器 x 的动作空间可以表示为：

$$A_x(t)=\{\Delta h_x^y(t_{next}),y\in Y\} \tag{11}$$

计算资源量每次的分配比例我们希望根据前一步动作的好坏来动态调整，如果上步动作大大缩短了服务响应时延，则增大一下次分配给该服务的计算资源量，具体由以下实现：

$$\Delta t=\left[C(S_t)-C(S_{t+1}\mid S_t,A_t)\right] \tag{12}$$

$$\Delta h_x^y(t_{next})=P\frac{2}{\Pi}\arctan\left(\frac{2\sqrt{3}}{\Pi}\arctan\Delta t\right) \tag{13}$$

其中 $C(S_t)$ 表示在状态 S_t 时的服务平均响应时延， $C(S_{t+1}|S_t,A_t)$ 表示在状态 $C(S_t)$ 执行动作 A_t 以后得到状态 S_{t+1} 时的服务平均响应时延，公式（12）表示时隙 t 与时隙 $t+1$ 的服务执行的平均响应时延差， Δt 越大说明执行动作 A_t 缩短的时延越大，则需要放大下一次的分配比例，反之则缩小下一次的分配比例，由于 $0\le\Delta t$ ，我们通过公式（13）来实现这一目的，式中 P 表示初始分配计算资源比例，该式通过反正切函数两次映射，能够将下一次的分配比例大小限制在 $(0,P)$ 。

在奖励函数的设置上，由于奖励值的大小跟动作执行的价值有关，因此当动作执行没有减小服务平均响应时延时给予该动作负值，否则奖励正值，奖励函数可以表示为：

$$R(t)=\begin{cases}-0.1 & \Delta t=0\\ \dfrac{\Delta w}{\alpha} & \Delta t>0\end{cases} \tag{14}$$

其中 α 为常数。

由于我们的目标是使得服务平均响应时延尽量小，相对的在算法中表示为最大化奖励值，这要求每一步动作都是有价值的，因此对于时间步长 T 的动作价值函数可以表示为未来奖励的总和：

$$Q(s,a)=\mathrm{E}\left(\sum_{t=1}^{T}\gamma R_t\mid S_t=s,A_t=a\right) \tag{15}$$

式中 $\gamma\in[0,1]$ 表示折扣因子， E 表示接下来 T 时间步长的价值期望，此时可以将最小化服务平均响应时延问题转变为寻找算法最大化动作值问题：

$$K = \arg\max_{a \in A} Q(s,a) \tag{16}$$

此外，为了让算法随着训练次数增大服务平均响应时延收敛到最优值，损失函数设置如下：

$$L(\theta_t) = \frac{1}{|D|}\sum_{j=1}^{D}\left(R_j + \gamma \max_{a'} Q(s',a';\bar{\theta}) - Q(s,a;\theta)\right)^2 \tag{17}$$

其中 s' 表示在状态 s 采取动作 a 以后得到的状态，D 表示用于存放执行每步动作以后训练元组 (s,a,R,s') 的经验池，参数 θ 表示神经网络权值，用梯度下降法更新：

$$\theta_{t+1} = \theta_t - \eta \nabla L(\theta_t) \tag{18}$$

其中 η 是学习率，在一定步数之后，更新参数 θ。

3 实验设置

3.1 实验参数与结果分析

表 1 仿真各种数据参数

参数名称	数据或表示
边缘服务器数量 S	4
服务请求数量 K	8
服务 k 数据转发量 d_k	[2, 10]GB
服务 k 所需计算资源 m_k	[2, 6]gigacyles
边缘服务器 i 计算能力 M_i	[4, 9]GHZ
边缘服务器间传输速率 v_{bet}	200Mbps
贪婪策略参数 ε 初始值	0.85
贪婪策略参数 ε 衰减值	0.0004
折扣因子 γ	0.85
学习率 η	0.0002

为了验证算法的有效性，我们将 DQN 算法与其他三种基准算法做了收敛性对比：基于服务请求数量的资源分配方法，基于贪心策略的资源分配方法和随机资源分配方法。

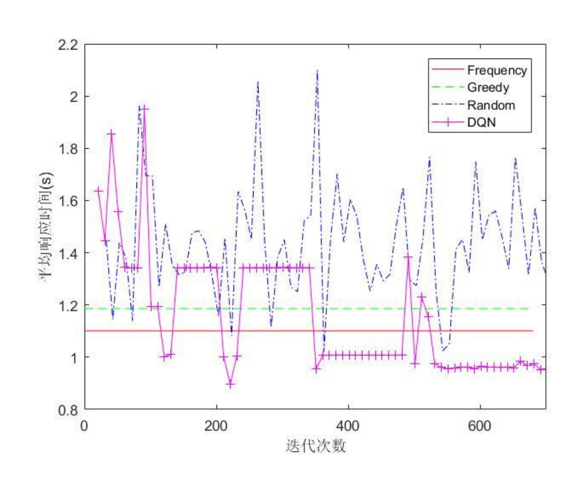


图 2 不同算法时延收敛对比图

从图 2 中可以看出，随着 DQN 算法迭代次数的增加，训练效果越来越好，最终在 600 次迭代之前收敛到最优值，服务平均响应时延从训练初期的 1.6s 缩小到不到 1s，值得注意的是，DQN 算法收敛值要优于其他基准算法，这说明深度强化学习算法在资源利用上有更加的

表现。

4 结论

本文研究了移动边缘计算环境中计算资源的分配问题，提出一种基于深度强化学习的移动边缘计算资源分配方法。该方法设计了深度强化学习中每一步动作分配计算资源量大小，对服务执行平均响应时延做出优化，实验表明，该方法能充分利用计算，并能获得更低的服务时延。

参 考 文 献:

[1] Zhang J, Chen B, Zhao Y, et al. Data security and privacy-preserving in edge computing paradigm: Survey and open issues[J]. IEEE access, 2018, 6: 18209-18237.

[2] 黄永明,郑冲,张征明等.大规模无线通信网络移动边缘计算和缓存研究[J].通信学报, 2021, 42(4): 44-61.

[3] Morabito I G. The internet of things: A survey[J]. Computer Networks, 2010.

[4] Pham Q. V., FangF, Ha V. N., et al. A Survey of Multi-Access Edge Computing in 5G and Beyond: Fundamentals, Technology Integration, and State-of the Ar. IEEE Access, 2020,8:1 16974-117017.

[5] 梁广俊,王群,辛建芳,李梦,许威.移动边缘计算资源分配综述[J].信息安全学报,2021,6(03):227-256.

[6] Zhang K, Mao Y M, Leng S P, et al. Energy-Efficient Ofloading for Mobile Edge Computing in 5G Heterogeneous Networks[J].IEEE Access, 2016, 4: 5896-5907.

[7] You C S, Zeng Y, Zhang R, et al. Asynchronous Mobile Edge Computation Offloading: Energy-Efficient Resource Management[J]. IEEE Transactions on Wireless Communications, 2018,17(11): 7590-7605.

[8] J. Huang, A. Zhou and S. Wang, "Price-Aware Service Deployment in Hierarchical Mobile-Edge Computing[J].in IEEE Internet of Things Journal, 2022,pp. 11533-11541.

(通讯作者: 黄昱泽 E-mail: huangyz@cqjtu.edu.cn)

作者贡献声明*:

黄昱泽：提出移动边缘计算资源分配的研究思路，确定研究命题；
冯北鹏：基于深度强化学习算法设计动作空间中反正切函数两次映射的资源分配函数，进行实验；
曹宇慧，郭珍珍：提供真实数据集和分析实验结果；
冯北鹏：论文起草；
黄昱泽：论文最终版本修订；